CDFSL-V: Cross Domain Few-Shot Learning for Videos

UCF Center for Research in Computer Vision
ICCV23 Paris

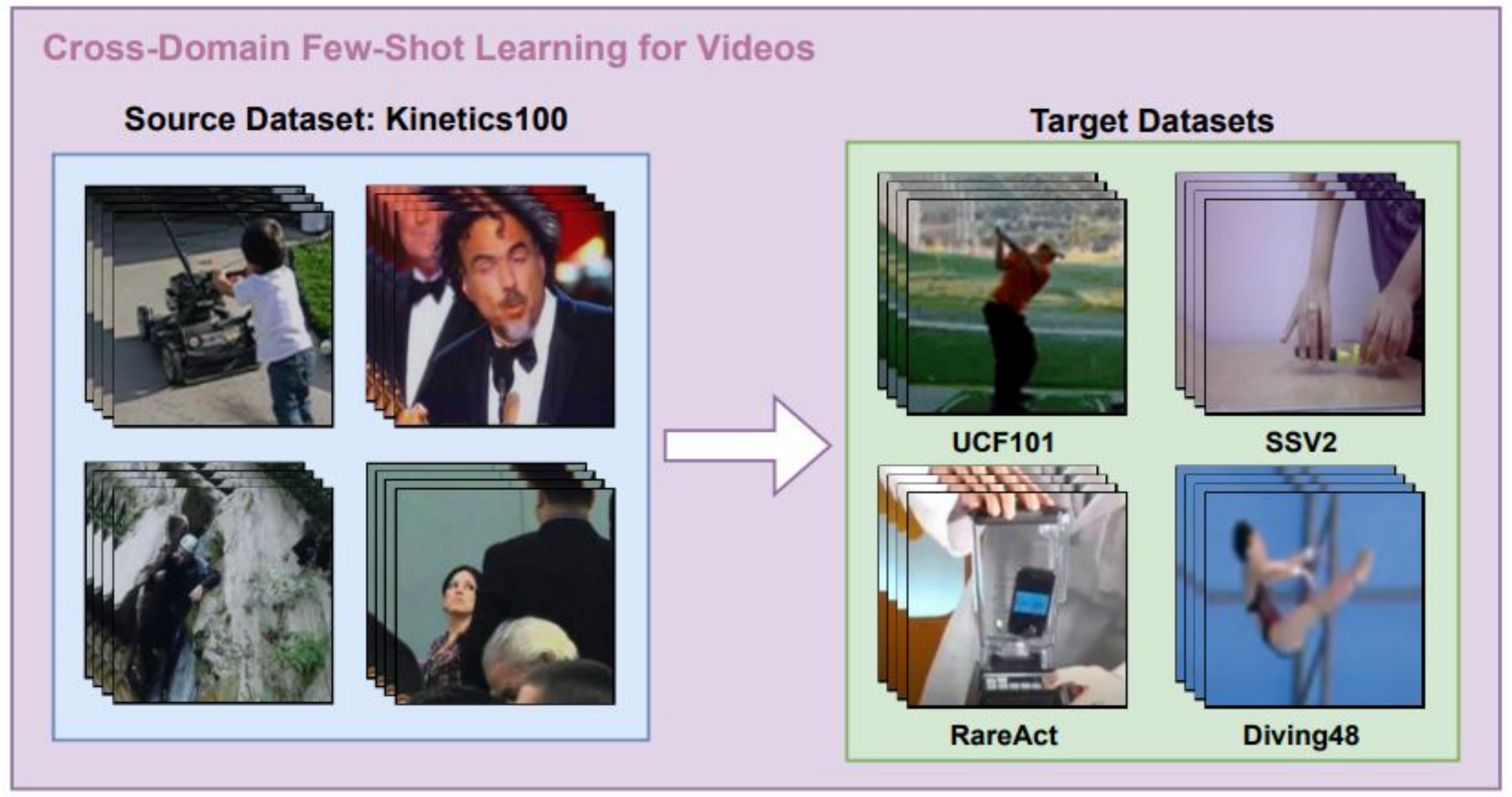
Sarinda Samarasinghe — sarinda.samarasinghe@ucf.edu
Mamshad Nayeem Rizve — mamshadnayeem.rizve@ucf.edu
Navid Kardan — nkardan@cs.ucf.edu
Mubarak Shah — shah@crcv.edu

Project Page:

Problem & Motivation

- Previously unexplored in the video setting
- Existing benchmarks overlap source/target classes
- We introduce a new evaluation protocol for CDFSL-V
- CDFSL-V outperforms existing image-based cross domain and video-based Few-shot techniques

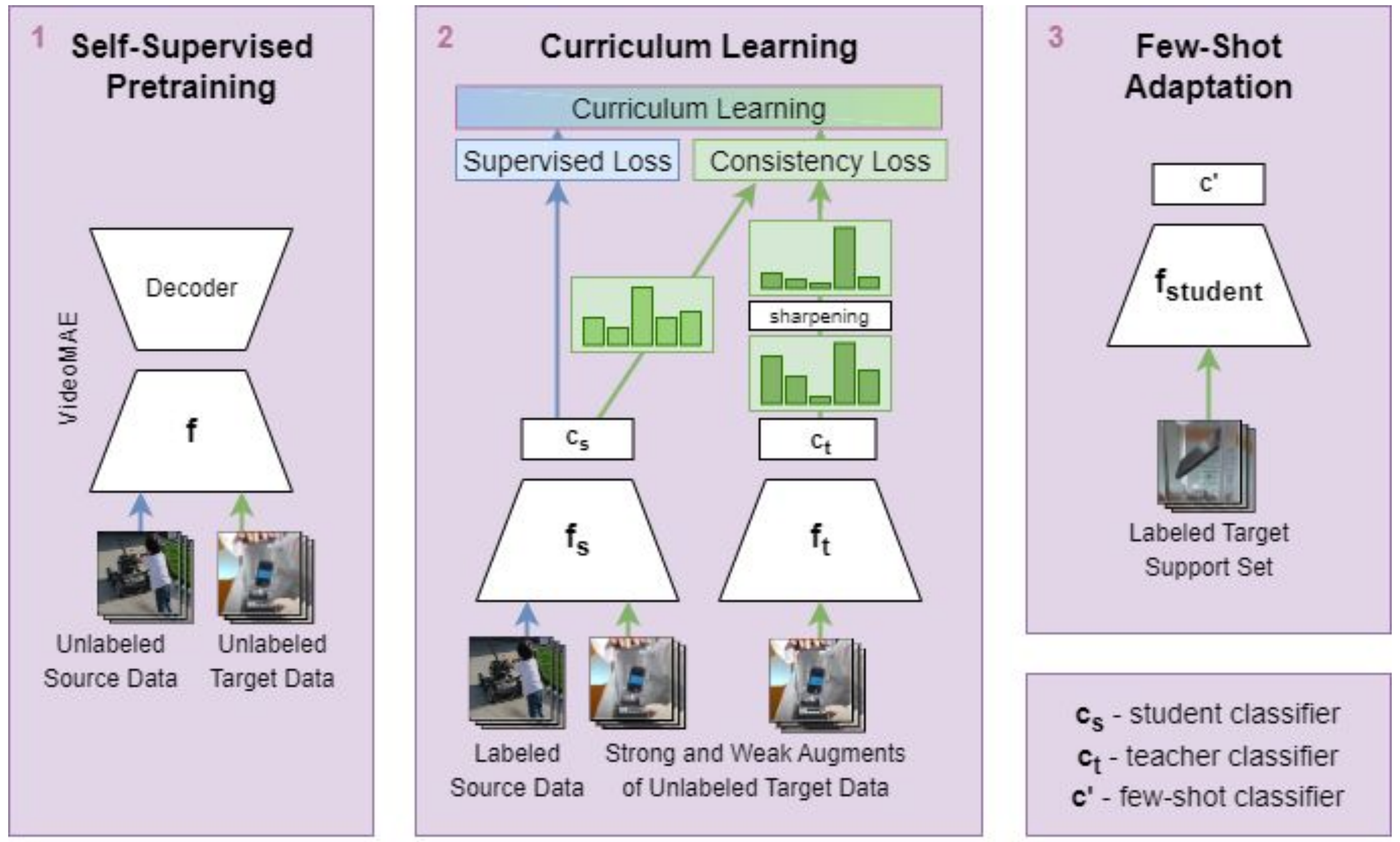
Our New Evaluation Protocol



Our Solution

- Self-supervised pre-training facilitates transferring of representations between domains
- Curriculum learning balances knowledge from the source and target dataset

Methodology



$$\mathcal{L}_{sup} = \mathcal{L}_{CE}(\text{Softmax}(f_s(\mathbf{x}_i)), \mathbf{y}_i), \quad \mathcal{L}_{con} = -\sum \hat{\mathbf{Y}} \log(\text{Softmax}(f_s(\mathbf{X}_{str}))),$$

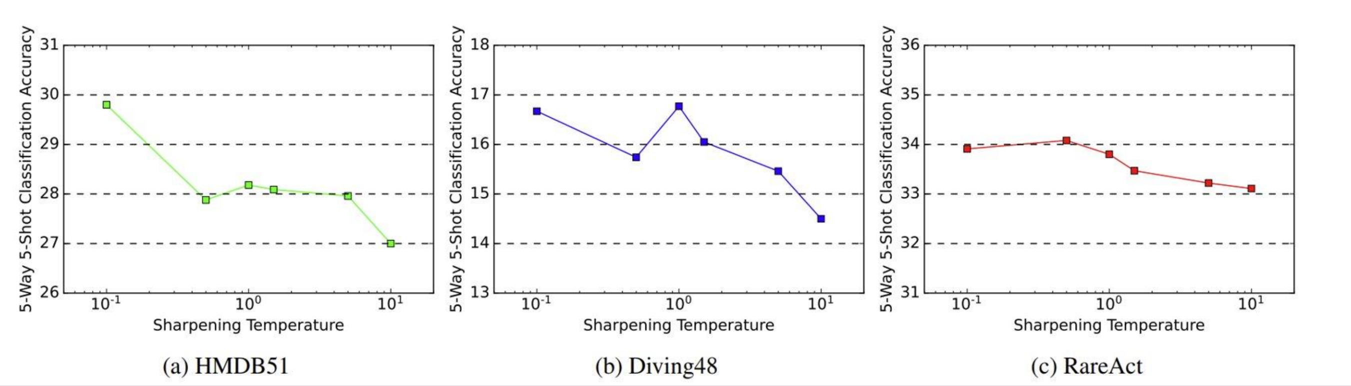
$$\mathcal{L}_{total} = \mathcal{L}_{sup} + \lambda \mathcal{L}_{con}$$

Ablations

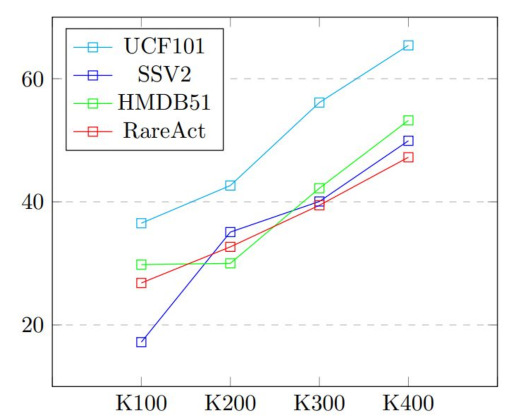
Principal Component Analysis

Method, Source Dataset: Kinetics-100	UCF101	HMDB51	SSV2	Diving48	RareAct	Average
Equal Loss Weighting	32.02	27.39	15.34	16.07	33.67	24.90
No Temperature Sharpening	34.01	28.18	15.21	16.77	33.80	25.59
Self-Supervised Training	37.54	25.09	16.21	17.14	29.58	25.11
Supervised Training	32.06	23.86	14.40	16.16	31.15	23.53
Ours	**36.53**	**29.80**	**17.21**	**16.37**	**33.91**	**26.82**

Effect of Temperature Sharpening



Vary Source Dataset Size



Experimental Results

5-Way 5-Shot Accuracy with Kinetics-400 as Source Dataset

Method	UCF101	SSV2	HMDB51	Diving48	RareAct	Average
Random Initialization	23.83	16.02	12.08	15.37	16.57	16.78
STARTUP++	60.82	39.60	44.71	14.92	45.22	41.05
Dynamic Distillation++	63.26	44.50	48.04	16.23	47.01	43.81
STRM	42.33	35.01	24.98	16.69	39.01	31.60
HYRSM	45.65	40.09	29.81	17.57	44.27	35.49
Ours	**65.42**	**49.92**	**53.23**	**17.84**	**49.80**	**47.24**

5-Way 5-Shot Accuracy with UCF101 as Source Dataset

Method, Source Dataset: UCF101	HMDB51	DIVING48	RareAct	Average
Random Initialization	21.69	14.48	26.98	21.05
STARTUP	23.56	14.84	31.31	23.24
Dynamic Distillation	24.06	**16.15**	32.00	24.07
Ours	**28.86**	16.07	**33.91**	**26.82**

5-Way 5-Shot Accuracy with Kinetics-100 as Source Dataset

Method, Source Dataset: Kinetics-100	UCF101	HMDB51	SSV2	Diving48	RareAct	Average
Random Initialization	23.83	16.02	12.08	15.37	16.57	16.78
STARTUP	32.20	24.97	15.16	14.55	31.77	23.73
Dynamic Distillation	34.10	25.99	16.00	16.24	31.20	24.71
Ours	**36.53**	**29.80**	**17.21**	**16.37**	**33.91**	**26.82**

5-Way 1-Shot Accuracy with Kinetics-100 as Source Dataset

Method, Source Dataset: Kinetics-100	UCF101	HMDB51	SSV2	Diving48	RareAct	Average
Random Initialization	23.83	16.02	12.08	15.37	16.57	16.78
STARTUP	24.48	16.66	14.17	13.13	17.21	17.13
Dynamic Distillation	26.04	17.44	14.96	13.73	19.02	18.24
Ours	**27.78**	**18.59**	**16.01**	**14.11**	**20.06**	**19.31**

5-Way 20-Shot Accuracy with Kinetics-100 as Source Dataset

Method, Source Dataset: Kinetics-100	UCF101	HMDB51	SSV2	Diving48	RareAct	Average
Random Initialization	32.33	27.97	15.12	15.83	33.53	24.96
STARTUP	34.02	30.48	17.15	17.30	38.45	27.48
Dynamic Distillation	36.72	33.09	17.56	17.33	39.97	28.93
Ours	**39.92**	**36.89**	**18.72**	**17.81**	**42.51**	**31.17**